Statistics and Computing (2026) 36:152
https://doi.org/10.1007/s11222-026-10903-y

ORIGINAL PAPER

An interpretable varying coefficients approach to non-linear regression

Davide Fabbrico[1] · Matteo Pedone[1] · Francesco Claudio Stingo[1]

Received: 11 July 2025 / Accepted: 12 May 2026

Abstract

Non-linear regression models are flexible approaches used to model complex associations. In many recent proposals, additional flexibility comes at the cost of loss of interpretability of the model's parameters and, consequently, of the data analysis results. This paper introduces a flexible model whose parameters are easily interpretable. In particular, the model incorporates non-linear effects through a semi-parametric spline-based representation that separates linear and non-linear effects via an orthogonal basis decomposition. We introduce a covariate-dependent regression coefficient to enhance flexibility and show the proposed approach's equivalence with a non-linear interaction model. In the proposed approach, the order of the covariates is relevant; however, we demonstrate that the model is invariant to this ordering. The proposed model performs comparatively well in simulation studies compared to state-of-the-art approaches. Finally, we illustrate the practical utility of the proposed approach through two applications that show a varying degree of non-linear associations.

Keywords Bayesian hierarchical modeling · Bayesian variable selection · Spline regression · Varying coefficients models

1 Introduction

Regression models with interaction effects can effectively model non-additive associations between the response variable and the covariates. These flexible regression models are commonly used in many fields, including ecology (Papadogeorgou et al. 2023), environmental science (Antonelli et al. 2020; Ferrari and Dunson 2020), and biomedical sciences (Cordell 2009; Im et al. 2023; Marchini et al. 2005). This paper introduces a Bayesian regression model designed for high-dimensional non-linear regressions. Building upon varying coefficient models (Hastie and Tibshirani 1993, VCM), we let covariates influence the outcome non-linearly, both as predictors and as effect modifiers. Consequently, the regression coefficients are smooth functions of the covariates. This dual role results in non-linear pairwise interaction terms, making the model flexible and interpretable. In summary, our method aims to: (i) select main and interaction effects; (ii) disentangle the functional form of associations between covariates/interaction terms and the response variable, and (iii) be defined by easily interpretable parameters.

There is extensive statistical literature on high-dimensional regression models that include linear and non-linear interaction terms; these models are typically paired with inferential approaches that result in the selection of the relevant effects. In particular, penalized regression approaches based on Lasso and Elastic Net have recently been proposed to account for two-way interactions in high-dimensional settings (Bien et al. 2013; Lim and Hastie 2015; Zhao et al. 2009). Although these methods provide interpretability in terms of main and pairwise interaction effects, they are often restricted to the linear regression model. Radchenko and James (2010) overcome this limitation by including either linear main effects and interaction or non-linear main effects and interactions. However, it is necessary to specify the functional form prior to the analysis, making their approach unable to automatically choose the most suitable functional form from the data, unlike our method.

Non-parametric methods, such as tree-based approaches (Chipman et al. 2010; Lampa et al. 2014) and Kernel Machines (Liu et al. 2018), are often used to handle com-

✉ Davide Fabbrico
davide.fabbrico@unifi.it

Matteo Pedone
mattpedone7@gmail.com

Francesco Claudio Stingo
francescoclaudio.stingo@unifi.it

[1] Department of Statistics, University of Florence, Viale Giovanni Battista Morgagni, Florence 50134, Italy

Published online: 29 May 2026

plex associations. However, these methods often sacrifice interpretability to create flexible models that yield accurate predictions. Several Bayesian approaches have been proposed to allow non-linearity and non-additivity. Savitsky et al. (2011) introduce spike-and-slab priors in the framework of Gaussian Processes (GP), developing a method that can account for both linear and non-linear associations in a parsimonious way at the cost of losing the interpretability of the selected terms. Building on Savitsky et al. (2011), Ferrari and Dunson (2020) uses a GP on top of a sparse linear regression interaction model to flexibly capture deviations from linearity. Ferrari and Dunson (2020) do not constrain the model to have a parametric form, and, if supported by the data, departures from linearity are accommodated through a GP. Although their approach yields an inference similar to ours, it treats non-linearity as a covariate-specific deviation from a linear effect rather than separating non-linear main effects from non-linear interaction terms as we do.

Other Bayesian spline-based methods, such as the approaches proposed by Antonelli et al. (2020) and Gustafson (2000), can induce sparsity and explicitly model interactions between covariates. However, these methods cannot simultaneously select variables and/or interaction terms along with their functional forms (linear vs. non-linear), a capability that distinguishes our approach.

Building on the Splines regression model of Scheipl et al. (2012), we propose a novel fully Bayesian approach, termed *Nonlinear Interactions Varying Coefficient* (INVENT) Regression Model. The key characteristics of our model can be outlined in four points: (i) *Non-linearity*. We allow main effects and pairwise interactions to have non-linear associations with the response variable through a spline-based representation. In particular, a suitable orthogonal basis decomposition allows the model to distinguish the functional form of the relationships (linear vs. non-linear). (ii) *Sparsity*. We induce sparsity in main effects and pairwise interactions through appropriate mixture priors, allowing for the selection of either linear or non-linear forms for main effects and interaction terms. As a result, (i) and (ii) yield a flexible semi-parametric model that retains the interpretability of the relationships between the response variable and the covariates. (iii) *Model representation guarantees*. We prove that using a spline-based representation of the covariates, both as predictors and effect modifiers, preserves the model's invariance to permutations in the covariate indices, thus overcoming the issue identified by Pedone et al. (2023). (iv) *Subject-specific inference*. A useful byproduct of our approach is that it includes subject-specific coefficients. Being smooth functions of the covariates, regression coefficients can capture a wide range of heterogeneous effects, borrowing strength among subjects with similar covariate patterns. Some of the aforementioned alternative methods share some characteristics with our approach; however, none of them encompasses all of these features simultaneously.

In high-dimensional settings with non-linear interactions, computational runtime becomes a critical consideration. Many existing Bayesian methods (*e.g.* Savitsky et al. 2011; Ferrari and Dunson 2020; Scheipl et al. 2012) rely on Monte Carlo Markov Chain (MCMC) sampling schemes that scale poorly with the number of covariates p and the number of observations n. This limitation arises primarily from two inherent challenges: the necessity to explore combinatorial interaction spaces, which introduce $\mathcal{O}(p^2)$ potential pairwise terms, and the requirement for numerically intensive non-conjugate updates to estimate non-linear effect structures.

In contrast, our proposed framework retains full model flexibility while resulting in improved computational efficiency in the evaluated settings. This is achieved without sacrificing interpretability, enabling efficient inference even for complex interaction models.

The paper is structured as follows. We introduce INVENT and discuss the prior distributions in Section 2. In Section 3, we present theoretical results that demonstrate the model's invariance to the ordering of covariates. We summarize the posterior inference in Section 4. Section 5 evaluates the finite sample performances through simulation studies, comparing our method with parametric and non-parametric models and state-of-the-art models for interaction estimation. In Section 6, we apply our method to various publicly available datasets to illustrate its inference and interpretability. Finally, Section 7 concludes the paper with a brief discussion.

2 Model construction

We introduce the model and its construction in Section 2.1. In Section 2.2, we present the semi-parametric spline-based approach used to model and select non-linear main effects and interactions. Finally, in Section 2.3, we discuss the prior distributions for all parameters and the covariate selection mechanism.

2.1 Bayesian varying-coefficient regression model

Let $y_i \in \mathbb{R}$ be the realization of a continuous random variable Y for subject i, for $i = 1, \dots, n$, where n is the sample size. Let $\boldsymbol{x}_i$ denote the row vector of observed values of p continuous covariates for subject i, such that $\boldsymbol{x}_i = (x_{i,1}, \dots, x_{i,p})$.

Regression models often assume that the effect of the covariates on the response is the same for all subjects. However, this assumption is not always realistic; we overcome this limitation by implementing subject-specific regression coefficients, which are smooth functions of the covariates. Building on the varying coefficients framework by Hastie and Tibshirani (1993), we define our model as:

$$y_i = \beta_0 + \sum_{j=1}^{p} f_j(\boldsymbol{x}_i) + \varepsilon_i \ , \tag{1}$$

where $f_j(\boldsymbol{x}_i)$ is a scalar-valued function defined as $f_j(\boldsymbol{x}_i) = \tilde{x}_{i,j}\, \beta_j(\tilde{\boldsymbol{x}}_i)$, with $\tilde{x}$ denoting a generic transformation of the covariates and $\beta_j(\tilde{x_i})$ representing the varying coefficient associated with the j-th covariate, which may depend on a subset of the covariates. The coefficient β_0 represents the intercept, and the error term ε_i follows a normal distribution $\mathcal{N}(0, \sigma^2)$. Note that if $\tilde{x}_{i,j} = x_{i,j}$ for all i and j and $\beta_j(\tilde{x_i}) = \beta_j$ for all j, meaning that the transformation is the identity and the regression coefficients remain constant across the population (i.e. independent of covariates), we recover the classical linear regression model; in Section 2.2 we will employ a basis function approach and we will have a vector $\tilde{x}_{ij}$ for each scalar $x_{i,j}$ and, in turn, $\beta_j(\cdot)$ will be a vector-valued varying coefficient.

In our framework, covariates serve a dual purpose since they enter Equation (1) as both predictors, with associated linear or non-linear effects, and modifiers of the effect of the other covariates, as described in Hastie and Tibshirani (1993). This dual role of covariates poses identifiability issues, as also noted in Pedone et al. (2023). In our model, we carefully construct the functional form of subject-specific coefficients to ensure identifiability in the likelihood function. For each covariate j, we constrain the set of effect modifiers to the parent set of j that results form an arbitrary ordering of the covaraites, defined as $pa(j) = \{h \mid h > j\}$; note that if we let each varying coefficient depend on every other covariate, we would obtain a model that is not identifiable, whereas our construction, in addition to being identifiable, enjoys the theoretical properties described in Section 3. We model these functions as:

$$\beta_j(\tilde{\boldsymbol{x}}_{i,pa(j)}) = \theta_j + \sum_{k \in pa(j)} \tilde{x}_{i,k}\, \omega_{j,k} \ , \tag{2}$$

where $\tilde{\boldsymbol{x}}_{i,pa(j)}$ denotes the observed values of the covariates indexed by $pa(j)$ for the i-th unit. Combining Equations (1) and (2), we obtain an interaction model that includes all possible two-way interactions among the transformed covariates $\{\tilde{\boldsymbol{x}}_j\}_{j=1}^p$.

The subject-specific coefficient in Equation (2) is defined by two sets of parameters: main effects $\{\theta_j\}_{j=1}^p$ and interaction terms $\{\omega_{j,k} \mid 1 \le j \le p-1,\ k \in pa(j)\}$. We assume that all transformations $\tilde{\boldsymbol{x}}_j$ are unknown smooth functions of the covariates. Given the dual role of the covariates, the transformed covariates $\{\tilde{\boldsymbol{x}}_j\}_{j=1}^p$ remain the same in both Equations (1) and (2). In the next section, we introduce a semi-parametric spline-based representation that enables the model to discern the functional form of the relationship (linear/non-linear) through an appropriate orthogonal basis decomposition. In the following, we use $\circ$ to represent the Hadamard product and $\cdot$ for the standard row-column matrix product.

2.2 Bayesian Orthogonal Basis P-Splines

Following Scheipl et al. (2012), we define $\boldsymbol{f}_j(\boldsymbol{x}_j) = \big(f_j(x_{1,j}),\ f_j(x_{2,j}),\ \ldots,\ f_j(x_{n,j})\big) \in \mathbb{R}^n$, as $\boldsymbol{f}_j(\boldsymbol{x}_j) = \tilde{\boldsymbol{x}}_j \cdot \boldsymbol{\beta}_j$, where $\tilde{\boldsymbol{x}}_j$ represents a suitable design matrix of B-spline bases for the covariate $\boldsymbol{x}_j$, and $\boldsymbol{\beta}_j$ is the corresponding spline coefficient. Note that $\boldsymbol{\beta}_j$ and $\beta_j(\cdot)$ have the same role but are not the same parameter; indeed, $\boldsymbol{\beta}_j$ are just working parameters that do not depend on the covariates, while $\beta_j(\cdot)$ are the actual coefficients of the proposed model. This transformation applies to all covariates, regardless of whether they are predictors or effect modifiers.

P-splines can be treated as a Bayesian hierarchical model (Ruppert et al. 2003); we assume $\boldsymbol{\beta} \mid s \sim \mathcal{N}(\boldsymbol{0}, s\boldsymbol{P}^{-1})$, where $\boldsymbol{P}$ is the singular penalty matrix constructed from the second-order differences of the adjacent spline coefficients, see Ruppert et al. (2003) for additional details, and s the prior variance. As noted in Scheipl et al. (2012), the coefficients associated with the constant and linear trends of all functions in the set $\{\boldsymbol{f}_j(\cdot)\}_{j=1}^p$ are not penalized, as they are in the null space of $\boldsymbol{P}$. We take a spectral decomposition of the variance-covariance matrix of $\tilde{\boldsymbol{x}}_j \cdot \boldsymbol{\beta}_j$,

$$Cov\left(\tilde{\boldsymbol{x}}_j \cdot \boldsymbol{\beta}_j\right) = s\, \tilde{\boldsymbol{x}}_j \cdot \boldsymbol{P}^{-1} \cdot \tilde{\boldsymbol{x}}_j^\top$$
$$= s\, \left(\boldsymbol{u}_j\ \boldsymbol{u}_0\right) \cdot \begin{pmatrix} \boldsymbol{d}_j & 0 \\ 0 & 0 \end{pmatrix} \cdot \left(\boldsymbol{u}_j\ \boldsymbol{u}_0\right)^\top \ ,$$

where $\boldsymbol{u}_j$ is the orthonormal matrix of eigenvectors with corresponding positive eigenvalues along the diagonal of the matrix $\boldsymbol{d}_j$, while $\boldsymbol{u}_0$ are the eigenvectors associated with the zero eigenvalues. In summary, we reparameterize and decompose each function as $\boldsymbol{f}_j(\boldsymbol{x}_j) = \boldsymbol{x}_j^\star \cdot \boldsymbol{\beta}_j^\star + \boldsymbol{x}_j^0 \beta_j^0$, where $\boldsymbol{x}_j^\star = \boldsymbol{u}_j \cdot \boldsymbol{d}_j^{1/2}$ is the orthogonal basis. In particular, $\boldsymbol{\beta}_j^\star$, which represents the non-linear effect, is a D_j-dimensional vector, where D_j denotes the number of eigenvectors and eigenvalues that explain most of the variability of $\boldsymbol{f}_j(\boldsymbol{x}_j)$, see Section 5.2 for additional details. The parameter $\beta_j^0 \in \mathbb{R}$ is the coefficient associated with the linear term $\boldsymbol{x}_j^0 \in \mathbb{R}^{n \times 1}$. This transformation is applied to all covariates; this modeling assumption, as detailed in Section 3, is needed to overcome the limitations encountered in Pedone et al. (2023) regarding the influence of covariate ordering that follows from the parent sets $pa(j)$.

Finally, assuming that the intercept terms are merged into the global intercept η_0 and the regression coefficients depend on the covariates, we can rewrite Equation (1) as:

$$y_i = \eta_0 + \sum_{j=1}^{p} \left[\boldsymbol{x}^\star_{i,j} \circ \boldsymbol{\beta}^\star_j(\boldsymbol{x}^\star_{i,pa(j)}) \right] \cdot \mathbf{1}_{D_j \times 1}$$
$$+ \sum_{j=1}^{p} x^0_{i,j}\, \beta^0_j(\boldsymbol{x}^0_{i,pa(j)}) + \varepsilon_i \ , \quad (3)$$

where $\mathbf{1}_{D_j \times 1}$ denotes a column vector of ones of dimension D_j, $\boldsymbol{x}^\star_{i,j}$ a D_j-dimensional vector representing the i-th row of the spline matrix corresponding to the j-th covariate, and $x^0_{i,j}$ is similarly defined analogously as a scalar. As the regression coefficients vary with the covariates, $\boldsymbol{\beta}^\star_j(\boldsymbol{x}^\star_{i,pa(j)})$ denotes a row vector of dimension D_j, while $\beta^0_j(\boldsymbol{x}^0_{i,pa(j)})$ is a scalar. Note that this representation allows us to disentangle the linear and non-linear components of each function $\boldsymbol{f}_j(\cdot)$ in Equation (1).

2.3 Parameter Expansion and Covariate Selection Priors

We build upon the parameter expansion of Normal Mixture of Inverse Gamma (peNMIG) prior structure of Scheipl et al. (2012), and define a covariate-dependent version. Specifically, we refer to this hierarchical structure, described in Figure 1, as peNMIGx; note that if the set $pa(j)$ is empty, the varying coefficient does not depend on any covariate, and the peNMIGx coincides with peNMIG structure.

The proposed peNMIGx is implemented for both the linear effects $\beta^0(\cdot)$ and the non-linear effects $\boldsymbol{\beta}^\star(\cdot)$. For the sake of presentation, we will focus on the non-linear ones, along with the corresponding design matrix of the spline bases $\boldsymbol{x}^\star$, and omit the superscript $\star$ for notation convenience.

Firstly, we expand the parameter $\boldsymbol{\beta}_j(\cdot) = \boldsymbol{\xi}_j\, \alpha_j(\cdot)$ to be the product of a scalar covariate-dependent $\alpha_j(\cdot)$ and a D_j-dimensional vector $\boldsymbol{\xi}_j$ that distributes $\alpha_j(\cdot)$ across the entries in $\boldsymbol{\beta}_j(\cdot)$. This parameter expansion technique, proposed by Gelman et al. (2008), improves the mixing of the MCMC algorithm, which has also been observed by Scheipl et al. (2012) and Ni et al. (2019). For the scalar $\alpha_j(\cdot)$, in accordance with Equation (2), we have proposed the following functional expressions:

$$\alpha_j(\boldsymbol{x}_{i,pa(j)}) = \theta_j + \sum_{k \in pa(j)} \boldsymbol{x}_{i,k} \cdot \boldsymbol{\omega}_{j,k} \ .$$

Therefore, following the reparametrization, our varying coefficients in Equation (2) can be rewritten as follows:

$$\boldsymbol{\beta}_j(\boldsymbol{x}_{i,pa(j)}) = \boldsymbol{\xi}_j \left(\theta_j + \sum_{k \in pa(j)} \boldsymbol{x}_{i,k} \cdot \boldsymbol{\omega}_{j,k} \right) \ . \quad (4)$$

Equations (3) and (4) yield an interaction model with main effect represented by the set $\{\theta_j\}_{j=1}^{p}$ and interaction effects

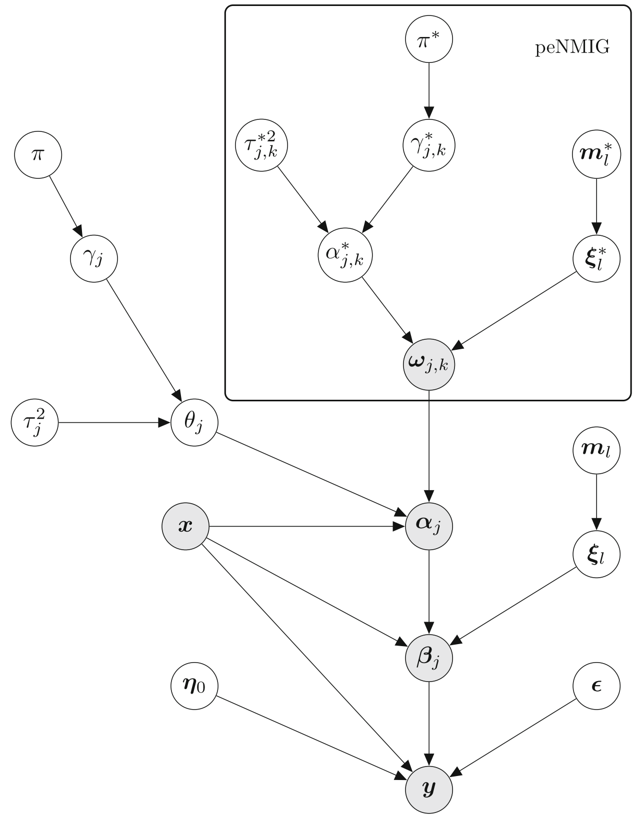

Fig. 1 DAG of the proposed model: the white nodes are the stochastic parameters and the grey nodes are the deterministic parameters. The superscript $*$ denotes the parameters of the peNMIG structure for the set of interaction effects $\{\boldsymbol{\omega}\}$, highlighted by the rectangle. To obtain a clearer representation, the priors for the linear and the non-linear parts of main effects ($\{\theta^0_j\}_j$ and $\{\theta^\star_j\}_j$) and interaction terms ($\{\omega^0_{j,k}\}_{j,k}$ and $\{\boldsymbol{\omega}^\star_{j,k}\}_{j,k}$) are jointly represented. In the graph, we omit the hyperparameters

represented by the set of column vectors $\{\boldsymbol{\omega}_{j,k}\}_{j,k}$. We consider scenarios where only certain covariates/interactions can influence the response.

Therefore, to introduce sparsity in the main effects $\{\theta_j\}_{j=1}^{p}$, we assign a spike-and-slab prior on the variance term of $\theta_j \sim \mathcal{N}(0, \lambda_j)$. Specifically, λ_j is parameterized as the product of a variance component, τ_j^2, and a binary latent variable γ_j, which determines whether the j-th variable is included or not in the model. For each γ_j we impose a discrete mixture distribution $\gamma_j \sim \pi\delta_1(\gamma_j) + (1-\pi)\delta_{\upsilon_0}(\gamma_j)$, where δ_1 is a Dirac delta function at 1. For the variance τ_j^2 we use a Inverse-Gamma distribution with parameters a_τ, b_τ, i.e. $\mathcal{IG}(a_\tau, b_\tau)$. The inclusion parameter γ_j takes the value 1 with probability π or a very small value υ_0 with probability $1-\pi$. For the weights π, we propose a $Beta(a_\pi, b_\pi)$ that provides an automatic adjustment for multiplicity, as described in Scott and Berger (2010). The entries of the vector $\boldsymbol{\xi}_j$ are assigned the following prior distribution:

$$\xi_{j,l} \sim \mathcal{N}(m_{j,l}, 1) \text{ , and } \quad m_{j,l} \sim \frac{1}{2}\delta_1(m_{j,l}) + \frac{1}{2}\delta_{-1}(m_{j,l}) \text{ ,} \tag{5}$$

where l is the index for the spline bases of covariate j, i.e. $l = 1, \ldots, D_j$. Finally, for the intercept η_0 we propose a normal distribution $\eta_0 \sim \mathcal{N}(0, 1)$ and for the variance of the error term $\sigma^2 \sim \mathcal{IG}(a_\sigma, b_\sigma)$.

The parameters $\boldsymbol{\omega}^\star_{j,k}$ and $\omega^0_{j,k}$, which represent the non-linear and linear interaction effects, respectively, between covariates j and k, are constructed following a peNMIG structure, as illustrated in Figure 1. Accordingly, the interaction effects are defined as $\boldsymbol{\omega}_{j,k} = \alpha^*_{j,k}\, \boldsymbol{\xi}^*_{j,k}$ where the sets $\{\alpha^*_{j,k}\}_{j,k}$ and $\{\boldsymbol{\xi}^*_{j,k}\}_{j,k}$ reflect a form of parameter expansion, as described for the varying coefficients $\{\boldsymbol{\beta}_j(\cdot)\}_j$. The parameters $\alpha^*_{j,k}$ and $\boldsymbol{\xi}^*_{j,k}$ follow the same prior structure as θ_j and $\boldsymbol{\xi}_j$, respectively. It is important to note that the parameters that govern the interaction effects are distinct from those associated with the main effects.

In summary, by constructing the peNMIGx prior structure, we are able to select both linear and non-linear main effects as well as linear and non-linear interaction effects.

3 Theoretical properties

The model defined in the previous sections depends on the parent sets $pa(j)$, $j = 1, \ldots, p$, which are determined by an arbitrary ordering of the covariates. Intuitively, for any chosen permutation, the parent set of variable j consists of all indices appearing after j in the ordering. As a result, different permutations of the design matrix can lead to different parent sets and, consequently, to apparently different interaction effects. In this section, we prove that the design matrix of the proposed model is invariant to the permutation of the covariates ordering, i.e. the design matrix remains identical regardless of the chosen permutation and all interaction effects remain the same.

As we mentioned earlier, in our approach, the same set of covariates serves two purposes: firstly, they contribute to Equation (1) as linear predictors with corresponding main effects; secondly, they function as effect modifiers (Hastie and Tibshirani 1993). This dual role results in two-way interaction effects that would make the model unidentifiable if all covariates had taken both roles. For instance, terms such as $\beta_j(x_k)x_j$ and $\beta_k(x_j)x_k$ represent the same interaction effect $x_j x_k$. The proposed parent set structure resolves this inherent non-identifiability by assigning the interaction unique to one parent set based on the ordering, restricting the set of effect modifiers to the parent set of each covariate $pa(j)$, as described in Pedone et al. (2023). Contrary to Pedone et al. (2023), which takes the reparameterization proposed by Scheipl et al. (2012) exclusively for the effect modifiers, we apply the same reparameterization to the covariates acting as linear predictors. In this way, we extend the model proposed by Pedone et al. (2023), achieving covariates order invariance.

In the following, we provide a formal definition of the problem, and Theorem 1 proves that our model is invariant under permutations of the ordering of the covariates. By construction, our linear predictor consists of two distinct parts: a linear one and a non-linear one. We focus on the non-linear component and consider continuous covariates only, since linear effects and categorical covariates are simpler, special cases. For clarity and without loss of generality, we assume that $D_j = D$ for all j, which implies an equal number of basis functions for each covariate.

Definition 1 *(Spline design matrix)* Let p denote the number of covariates, and n the number of observations. For a given ordering of the indices of the covariates $\{1, \ldots, p\}$, the matrix $\boldsymbol{X} = \left(\boldsymbol{x}^\star_1, \ldots, \boldsymbol{x}^\star_p\right) \in \mathbb{R}^{n \times pD}$ represents the *Spline design matrix* of the model that includes non-linear effects only, where each element is an orthogonal basis matrix.

Definition 2 *(Permutation set)* Let σ be an arbitrary permutation. We define the *Permutation set*, denoted as $\mathcal{C}^\sigma$, as the set formed by all ordered pairs of combinations from the image set of the function σ, $Im(\sigma)$. Formally,

$$\mathcal{C}^\sigma = \left\{(j, k) \mid j \in Im(\sigma), k \in pa^\sigma(j)\right\} \text{ .}$$

where $pa^\sigma(j)$ denotes the set of indices succeeding j under the permutation σ; for further details, please refer to Section 1.2 of the Supplementary Material.

Definition 3 *(Permutation interaction matrix)* Let σ be an arbitrary permutation. We define the *Permutation interaction matrix*, denoted by $\boldsymbol{\Sigma}^\sigma \in \mathbb{R}^{p \times pD}$, as the structure representing the interaction effects between covariates, according to the following rule:

$$\boldsymbol{\Sigma}^\sigma_{j,k} = \begin{cases} \boldsymbol{\omega}^\top_{j,k} & \text{if } (j, k) \in \mathcal{C}^\sigma \text{ ,} \\ \mathbf{0}_{1 \times D} & \text{if } (j, k) \notin \mathcal{C}^\sigma \text{ ,} \end{cases}$$

where $j, k = 1, \ldots, p$, and $\mathbf{0}_{1 \times D}$ denotes a D-dimensional zero row vector.

Let $\boldsymbol{\theta} = \left(\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_p\right)$ be a $n \times p$ matrix, where $\boldsymbol{\theta}_j = \theta_j \cdot \mathbf{1}_{n \times 1}$, $\boldsymbol{\Xi}$ is a block diagonal matrix with elements $\{\boldsymbol{\xi}^\top_j\}_{j=1}^p$ and $\mathbf{1}$ is a vector of ones of dimension $pD \times 1$.

Theorem 1 *Under any permutation σ, Equation (3) can be rewritten in vector form as:*

$$\boldsymbol{y} = \boldsymbol{\eta}_0 + \boldsymbol{X} \circ \left\{\left[\boldsymbol{X} \cdot \left(\boldsymbol{\Sigma}^\sigma\right)^\top + \boldsymbol{\theta}\right] \cdot \boldsymbol{\Xi}\right\} \cdot \mathbf{1} + \boldsymbol{\varepsilon} \text{ ,}$$

where the permutation interaction matrix $\boldsymbol{\Sigma}^{\sigma}$, *is the only element that depends on the permutation* σ.

Theorem 1 highlights that under the proposed model, any permutation σ of the covariates will produce an equation that depends on the same spline design matrix defined by the chosen arbitrary ordering of the covariates, i.e. that depends on the same $\boldsymbol{X}$ terms. Crucially, different permutations lead to an algebraically equivalent linear predictor, generating the same main effects and interactions. They differ only in the parameterization of the coefficients, meaning the parameters are structurally identical up to a relabeling.

Corollary 1 *The linear predictor on the right-hand side of Equation (3) associated with two different generic permutations,* σ *and* π, *possesses the same numerical value if* $\boldsymbol{\Sigma}^{\sigma} = \boldsymbol{\Sigma}^{\pi}$.

Corollary 1 states that if the parameters of two alternative orderings of the covariates coincide, then the linear predictors are identical. All proofs, a series of examples, and additional in-depth analyses of these theoretical properties are detailed in Section 1.3 of the Supplementary Material. Furthermore, in Section 1.3 we empirically assess the method's robustness to the ordering of the covariates, i.e. to alternative permutations of the design matrix columns.

4 Posterior computation

We implement an MCMC algorithm to obtain the posterior distribution of the parameters of interest. Posterior inference is carried out through a Metropolis-within-Gibbs sampler, as outlined in Algorithm 1. A detailed description of the algorithm can be found in Section 2 of the Supplementary Material.

The algorithm is initialized at a random point in the parameter space and then used to generate draws from the posterior distribution. After burn-in and thinning, inference is carried out on the remaining samples. The selection for main effects and interaction effects is based on the marginal posterior probability of inclusion (MPPI).

Algorithm 1: peNMIGx MCMC sampler under no heredity assumption

procedure SAMPLER
Initialize all the parameters in Figure 1
for *iterations* $t = 1, \dots, T$ **do**
  update π^*_{lin} and π^*_{nlin} from its *Full Conditional Distribution*
  for $j = 1, \dots, p$ **do**
    /* Update these parameters first to check the heredity conditions */
    update $\gamma_{j,lin}$ and $\gamma_{j,nlin}$ from its *Full Conditional Distribution*
  for $j = 1, \dots, p-1$ **do**
    for $k = j+1, \dots, p$ **do**
      update $\gamma^*_{j,k,lin}$ and $\gamma^*_{j,k,nlin}$ from its *Full Conditional Distribution*
      update $\tau^{*2}_{j,k,lin}$ and $\tau^{*2}_{j,k,nlin}$ from its *Full Conditional Distribution*
      update $\alpha^*_{j,k,lin}$ and $\alpha^*_{j,k,nlin}$ with a *Metropolis-Hastings* step
    for $k = j+1, \dots, p$ linear *(*$l = 1, \dots, D_k$ non-linear*)* **do**
      update $m^*_{j,k,lin}$ and $\boldsymbol{m}^*_{j,k,l,nlin}$ from its *Full Conditional Distribution*
      update $\xi^*_{j,k,lin}$ and $\boldsymbol{\xi}^*_{j,k,l,nlin}$ with a *Metropolis-Hastings* step
  rescale $\boldsymbol{\xi}^*$ and $\boldsymbol{\alpha}^*$
  compute $\omega_{j,k,lin}$ and $\boldsymbol{\omega}^*_{j,k,nlin}$
  update π_{lin} and π_{nlin} from its *Full Conditional Distribution*
  for $j = 1, \dots, p$ **do**
    update $\tau^2_{j,lin}$ and $\tau^2_{j,nlin}$ from its *Full Conditional Distribution*
    update $\theta_{j,lin}$ and $\theta_{j,nlin}$ with a *Metropolis-Hastings* step
  compute $\alpha_{j,lin}(\boldsymbol{x}_{pa(j)})$ and $\boldsymbol{\alpha}_{j,nlin}(\boldsymbol{x}_{pa(j)})$
  for $j = 1, \dots, p$ linear *(*$l = 1, \dots, D_j$ non-linear*)* **do**
    update $m_{j,lin}$ and $\boldsymbol{m}_{j,l,nlin}$ from its *Full Conditional Distribution*
    update $\xi_{j,lin}$ and $\boldsymbol{\xi}_{j,l,nlin}$ with a *Metropolis-Hastings* step
  rescale $\boldsymbol{\xi}$ and $\boldsymbol{\alpha}(\cdot)$
  compute $\beta_{j,lin}(\boldsymbol{x}_{pa(j)})$ and $\boldsymbol{\beta}_{j,nlin}(\boldsymbol{x}_{pa(j)})$
  update η_0 from its *Full Conditional Distribution*
  update σ^2 from its *Full Conditional Distribution*

Interaction effects may depend on the presence of the corresponding main effects. A common assumption made on two-way interactions is weak heredity, which states that an interaction term is included if at least one of the two main terms is included. As noted in Bien et al. (2013), the weak heredity approach acts as a robust compromise: it promotes parsimony and interpretability, while retaining the flexibility to identify interactions even when one main effect is negligible. We evaluate the inclusion of interaction terms conditional on the inclusion of corresponding main effects (see also Stingo et al. 2011). In the next section, we will present the results related to this specific assumption. In Section 3 of the Supplementary Material, results for both strong and no heredity assumptions will be provided; the strong heredity assumption requires that an interaction term can be selected only if both corresponding main terms are already included in the model, while no heredity imposes no constraints.

The implemented R package provides two core functions. The first function invMCMC runs a single MCMC chain, and the user can set the hyperparameter values, specify the heredity constraints between main effects and interactions, and other advanced options. Depending on the selected settings, the function returns either the full set of posterior samples or only the posterior means of the parameters, which is helpful for reducing memory usage when working with large datasets. The second function, invParMCMC, automates the parallel execution of multiple MCMC chains and computes convergence diagnostics such as $\hat{R}$ and the effective sample size (ESS). The package also includes advanced tools for posterior visualization (e.g. Marginal Posterior Probability of Inclusion plots and subject-specific plots), along with convergence diagnostics. Further implementation details can be found in the package referenced in the Supplementary Material section of this document.

5 Numerical experiments

In this section, we empirically investigate our model's finite sample performance in terms of prediction and covariate selection. Section 5.1 outlines the data generative mechanism. Section 5.2 elucidates and justifies the selection of hyperparameters for the proposed model. Finally, Section 5.3 presents the results in various scenarios. Notably, we benchmark our model against established parametric and non-parametric methods, including Bayesian Group Lasso (BLASSO, Xu and Ghosh 2015), NLinteraction (NLint) by Antonelli et al. (2020), Support Vector Machine (SVM) by Hearst et al. (1998), and Bayesian Additive Regression Tree (BART) by Chipman et al. (2010). In Section 5 of the Supplementary Material, we provide additional comparisons with spikeSlabGAM (Scheipl 2011), MixSelect (Ferrari and Dunson 2020), and VCBART (Deshpande et al. 2024) in scenarios with a limited number of covariates.

BLASSO revisits the Bayesian group lasso and uses spike-and-slab priors for group variable selection. As currently defined, this model does not allow interactions. However, it is possible to include interactions using the strategy proposed in Im et al. (2023). NLint is a Bayesian regression model that can identify non-linear main effects and interactions using a semi-parametric approach based on Splines. Additionally, it induces sparsity through the use of spike-and-slab priors. SVM and BART are well-established non-parametric techniques that typically perform well in terms of prediction but do not provide a direct interpretation of the effect of the covariates. As a unique feature, the proposed model disentangles the linear from the non-linear effects; this, along with the presence of spike-and-slab prior, enables us to identify the relevant variables specific to linear and/or non-linear effects. Furthermore, by construction, INVENT results in subject-specific regression coefficients rather than constant effects for all individuals.

5.1 Generating mechanism

In these simulation studies, we simulate n samples and p continuous covariates from a normal distribution with vector mean $\mathbf{0}$ and covariance matrix $\boldsymbol{I}_n$, where $\boldsymbol{I}_n$ is a $n \times n$ identity matrix. The global intercept is sampled from a uniform distribution, $\eta_0 \sim U(1, 2)$; we then changed the sign of the intercept with probability 0.5. The vector of error terms $\boldsymbol{\varepsilon}$ is generated from a multivariate normal distribution with mean vector $\mathbf{0}$ and covariance matrix $0.5\boldsymbol{I}_n$. Among the p covariates, a fixed number of main effects are randomly selected. For each selected covariate, the corresponding coefficient θ_j is drawn from a uniform distribution, $\theta_j \sim U(0.7, 1)$, with a randomly assigned sign. Specifically, linear and non-linear effects have the same sign. In the model, the main effects are either entirely linear or non-linear. A non-linear main effect always comprises both linear and non-linear components. Interaction effects are generated based on the heredity assumption. For strong heredity, we define $\omega_{j,k} = \theta_j \theta_k$. For weak or no heredity, interaction terms are randomly sampled from the corresponding candidate sets and their coefficients are drawn from $\mathcal{N}(2, 0.5)$ with a random sign. Under weak heredity, we ensure that at least one of the generating effects is non-zero. The same procedure applies to non-linear interaction terms, where $\boldsymbol{\omega}^{\star}_{j,k}$ are D-dimensional vectors with repeated entries. Finally, the elements of $\boldsymbol{\xi}$ in Equation (5) are sampled from $\mathcal{N}(m_{j,l}, 1)$, where $m_{j,l}$ are sampled according to Equation (5).

5.2 Hyperparameter settings

Hyperparameters a_π, b_π for linear and non-linear main effects and a_{π^*}, b_{π^*} for linear and non-linear interaction effects are set to induce a weakly informative Beta prior. In particular, we have considered $a. + b. = 2$ for both main effects and interaction effects, and the prior expected mean $m = a./(a. + b.)$ to a small value $m = 0.3$ for the linear main and interaction effect and $m = 0.1$ for the non-linear main and interaction effect, that corresponds to a 30% and 10% prior probability of inclusion of a linear effect and non-linear effect respectively. The remaining hyperparameters of the peNMIG priors, for both linear main and interaction and the non-linear main and interaction effects, are set following the guidelines provided by Scheipl et al. (2012): $v_0 = 0.00025$, $(a_\tau, b_\tau) = (5, 25)$. The hyperparameters for the variance for the error terms are set to $(a_\sigma, b_\sigma) = (3, 1)$. Following Scheipl et al. (2012), in order to gain computational efficiency, Splines bases are constructed using only the first few eigenvectors and eigenvalues that explain 95% of the variability of $\boldsymbol{f}_j(\boldsymbol{x})$. For more details on the data generating mechanism, hyperparameter settings, and robustness to hyperparameter settings, see Section 4 of the Supplementary Material.

5.3 Simulation scenarios

We designed four scenarios using data simulated from the proposed model. Scenario 1 features purely linear main and interaction effects. In Scenario 2, we combined linear main effects with non-linear interactions. Scenario 3 mixes non-linear main effects with linear interactions. Lastly, Scenario 4 incorporates both non-linear main and interaction effects. Each scenario involves the analysis of two settings, both with a sample size of $n = 500$, and differing only in the dimensionality of the covariate space: $p = 25$ and $p = 100$. The proportion of non-null main effects is set to 30% when $p = 25$, and to 5% when $p = 100$.

The magnitudes of all non-zero regression coefficients are constrained to lie within the interval [0.7, 1.0]. The results under the weak heredity assumption are summarized in Table 1.

To assess the performance in variable selection, comparisons are made in terms of True positive rate (TPR), False positive rate (FPR) and Matthews correlation coefficient (MCC) a measure of overall selection accuracy, that takes into account true and false positives and negatives (TP, FP, TN, FN, respectively)

$$TPR = \frac{TP}{TP + FN}\ , \quad FPR = \frac{TN}{TN + FP}\ ,$$
$$MCC = \frac{TP \cdot TN - FP \cdot FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}\ .$$

We calculate TPR and FPR stratified by effect type, linear vs. non-linear, and component, main effects vs. interactions. However, for simplicity, in Table 1 we report aggregated TPR and FPR metrics. Detailed stratified results, including performance splits for linear and non-linear effects, are provided in Section 3 of the Supplementary Material.

In this section, we report the results obtained using INVENT, NLInt, BLASSO, SVM, and BART. In Section 5 of the Supplementary Material we compare the proposed model with spikeSlabGAM, MixSelect, and VCBART.

In the NLint model, associations are modeled using splines without separation for linear and non-linear effects. Therefore, the calculation of linear and non-linear TPR and FPR is performed using the same set of variables selected by the model. In the BLASSO model, non-linear metrics are not calculated. For the INVENT, BLASSO and NLint models, covariates are selected if included in the median probability model (Barbieri and Berger 2004).

To evaluate the predictive capacity of the model, we calculate the out-of-sample Mean Squared Error (MSE); we randomly select 80% of the samples as the training set and then use the remaining 20% as the validation set. For INVENT and BLASSO, the out-of-sample MSE is computed based on the posterior predictive distribution. For the other methods, NLint, BART, and SVM, the default settings of the respective packages were used.

Table 1 shows that, for $p = 25$ and $p = 100$, our model performs well in all four scenarios, correctly identifying the non-null and null covariates generated and obtaining good out-of-sample prediction values. The NLint model also achieves good prediction results, but since it does not separate linear and non-linear effects, the corresponding FPR for the main effects does not compare well with respect to INVENT; consequently, our method also performs better in terms of MCC and MSE. The BLASSO model achieves good results in the linear scenario, but its performance decreases with the presence of non-linear effects in the data. On the other hand, BART is highly flexible and performs well in out-of-sample predictions. Finally, the SVM method does not show strong predictive capabilities.

While predictive accuracy and variable selection are central to our evaluation, the computational cost of MCMC in non-linear regression represents a critical practical constraint, particularly in the presence of interaction terms and complex hierarchical structures. To provide a complete analysis, Figure 2 compares the execution times of all methods under Scenario 4, with $n = 500$ and $p = 25$ on the left-hand side and with $p = 100$ on the right-hand side, setting the number of MCMC iterations to 1000 for all MCMC-based approaches. We complement these runtime analyses with out-of-sample MSE calculations to check an approximate convergence of the methods. Additional analyses are presented in Section 3 of the Supplementary Material.

In Figure 2, it can be observed that our model exhibits greater computational efficiency compared to MCMC-based approaches in the contexts analyzed. Specifically, with a fixed number of iterations for all methods, our model achieves better out-of-sample MSE performance compared to the considered alternatives. However, it is important to emphasize that fixing the number of iterations represents an intrinsic limitation: in practice, some methods might require fewer or more iterations to achieve convergence, potentially affecting the direct comparability of execution times. Thus, Figure 2 mainly aims to highlight differences in computational cost under equivalent operational conditions.

We evaluated the proposed method under varying sample sizes n and numbers of covariates p. When the number of covariates is large, the computational performance of our method is comparable to, and in some cases slightly better than, existing approaches in the literature, as illustrated in the right panel of Figure 2. In scenarios with large sample sizes ($n \geq 10{,}000$) and a moderate number of covariates ($p = 10$), our method achieves a substantial improvement in computational efficiency, being approximately 10 times faster than BART. All comparisons were conducted with the number of iterations fixed at 1000. Our method consistently achieved a lower out-of-sample MSE compared to BART.

Table 1 Simulation results for $p = 25$ and $p = 100$: means across 50 replications (standard errors in parentheses). In each scenario and for each index the best performance is in bold. The notation / indicates cases where variable selection is not performed.

Method	TPR	FPR	MCC	MSE
$\boldsymbol{p = 25}$				
Scenario 1				
INVENT	**0.893** (0.102)	**0.000** (0.000)	**0.943** (0.056)	**0.259** (0.038)
BLASSO	0.762 (0.076)	0.001 (0.002)	0.847 (0.061)	2.413 (0.869)
NLint	0.856 (0.098)	0.034 (0.006)	0.565 (0.038)	0.516 (0.271)
SVM	/	/	/	6.462 (3.140)
BART	/	/	/	2.636 (0.955)
Scenario 2				
INVENT	0.904 (0.074)	**0.000** (0.000)	**0.944** (0.044)	**0.339** (0.062)
BLASSO	0.388 (0.174)	0.005 (0.006)	0.522 (0.158)	6.476 (1.431)
NLint	**0.992** (0.026)	0.030 (0.004)	0.735 (0.021)	1.128 (0.441)
SVM	/	/	/	10.492 (3.246)
BART	/	/	/	8.572 (3.259)
Scenario 3				
INVENT	**0.946** (0.052)	**0.000** (0.000)	**0.969** (0.030)	**0.311** (0.042)
BLASSO	0.404 (0.083)	0.004 (0.002)	0.563 (0.077)	6.578 (1.428)
NLint	0.933 (0.049)	0.021 (0.006)	0.757 (0.029)	0.612 (0.128)
SVM	/	/	/	8.308 (1.963)
BART	/	/	/	2.379 (0.739)
Scenario 4				
INVENT	0.906 (0.071)	**0.000** (0.000)	**0.946** (0.036)	**0.341** (0.065)
BLASSO	0.241 (0.072)	0.006 (0.006)	0.391 (0.119)	10.323 (3.912)
NLint	**0.944** (0.012)	0.012 (0.004)	0.867 (0.036)	1.484 (0.532)
SVM	/	/	/	14.374 (5.449)
BART	/	/	/	9.634 (4.139)
$\boldsymbol{p = 100}$				
Scenario 1				
INVENT	0.670 (0.226)	**0.000** (0.000)	**0.800** (0.173)	**0.320** (0.064)
BLASSO	0.670 (0.157)	**0.000** (0.000)	0.796 (0.110)	2.715 (1.283)
NLint	**0.700** (0.287)	0.002 (0.001)	0.461 (0.136)	0.739 (1.227)
SVM	/	/	/	3.544 (2.017)
BART	/	/	/	1.810 (0.994)
Scenario 2				
INVENT	**0.691** (0.21)	**0.000** (0.000)	**0.819** (0.144)	**0.439** (0.187)
BLASSO	0.301 (0.161)	**0.000** (0.000)	0.483 (0.198)	5.462 (2.096)
NLint	0.620 (0.404)	0.001 (0.001)	0.483 (0.239)	2.723 (2.088)
SVM	/	/	/	6.374 (2.672)

Table 1 continued

Method	TPR	FPR	MCC	MSE
BART	/	/	/	5.600 (2.477)
Scenario 3				
INVENT	0.785 (0.124)	**0.000** (0.000)	**0.883** (0.072)	**0.349** (0.075)
BLASSO	0.367 (0.164)	**0.000** (0.000)	0.498 (0.195)	5.273 (1.705)
NLint	**0.920** (0.061)	0.001 (0.001)	0.712 (0.002)	0.497 (0.010)
SVM	/	/	/	5.953 (2.746)
BART	/	/	/	2.070 (1.237)
Scenario 4				
INVENT	0.677 (0.130)	**0.000** (0.000)	**0.819** (0.082)	**0.544** (0.178)
BLASSO	0.170 (0.089)	0.001 (0.001)	0.277 (0.160)	9.450 (3.349)
NLint	**0.920** (0.155)	0.001 (0.001)	0.781 (0.067)	1.309 (1.370)
SVM	/	/	/	7.256 (2.143)
BART	/	/	/	5.343 (2.162)

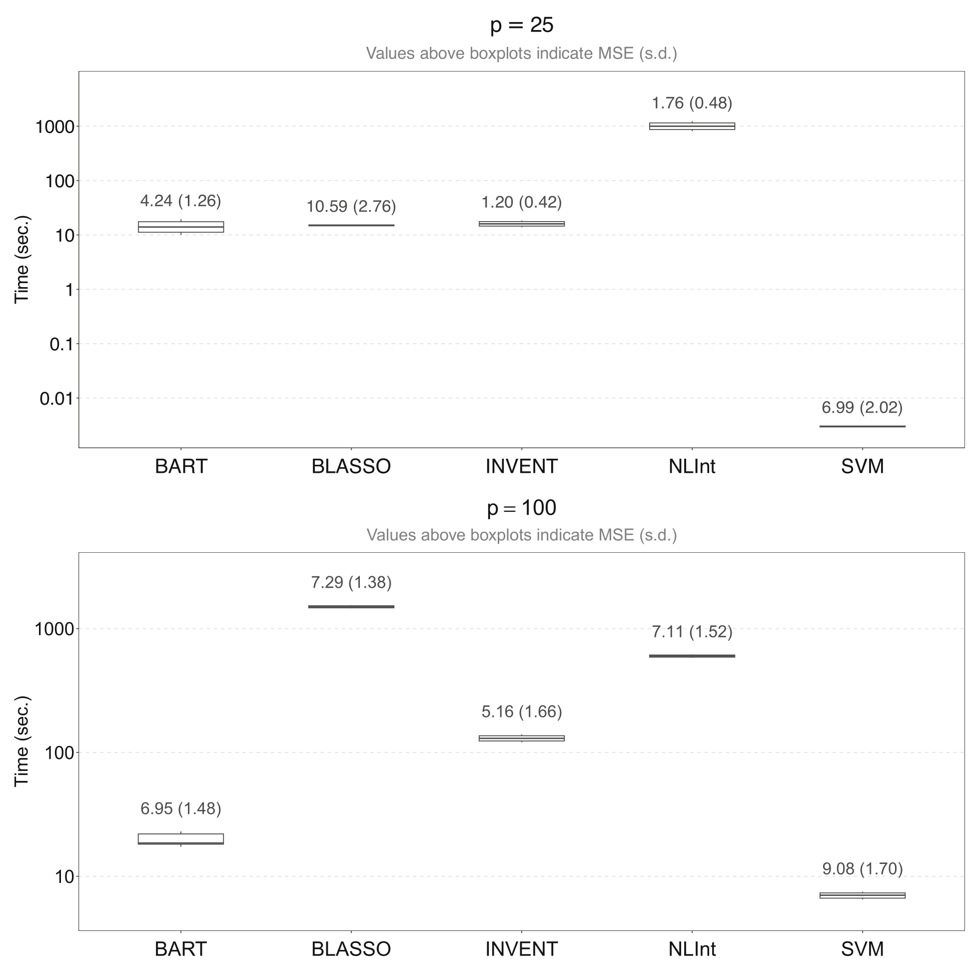

Fig. 2 Execution times in seconds (y-axis on a log-scale) and corresponding MSEs (with standard deviations in parentheses) across 20 replications for each method

More details can be found in Section 3 of the Supplementary Material.

6 Real data analysis

In this section, we illustrate the practical utility of the proposed approach through the analysis of two datasets in comparison with well-established and alternative approaches. We employ an 80-20 split, using 80% of the data for training and the remaining 20% for testing, and assess predictive performances by computing the mean squared error (MSE) on the test set. The analysis also includes comparison with the following alternative approaches: Multiple Linear Regression (MLR), Bayesian Group Lasso (BLASSO), Support Vector Machine (SVM), Bayesian Additive Regression Trees (BART), implemented using the package proposed by Chipman et al. (2016), NLinteraction (NLint), MixSelect, and VCBART.

The same hyperparameter and tuning parameter settings as those used in the simulation studies are employed. For VCBART, we adopted the hyperparameters described in Deshpande et al. (2024) and set the effect modifiers to coincide with the predictors. For INVENT, BLASSO, and NLint, variables included in the median probability model are selected.

We assume weak heredity on the relationship between main effects and interactions. In our analysis, discrete covariates are included in the model only through linear main effects or linear interactions.

6.1 Prediction of body mass index

The first application is based on real-world data obtained from the National Center for Health Statistics (NHANES) website. It consists of survey data collected between 1999 and 2006, providing information on the health of adults and children in the United States. The aim of the analysis is to predict BMI using laboratory data and demographic information. The laboratory data include, for example, creatine, glucose, and albumin, while the demographic information includes race, gender, age, and education level. After missing data are removed, the data set comprises 3937 observations and 28 covariates.

The MCMC chain was run for a total of 100,000 iterations, with the first 50,000 iterations discarded as burn-in and a thinning interval of 2 applied. The predictive results reported in the first column of Table 2 suggest a possible presence of non-linearity in the data. Non-linear models, such as INVENT, BART, and NLint, show comparable performance in terms of out-of-sample MSE, with a slight improvement over the linear approaches. We ran 5 parallel MCMC independent chains to promote stable convergence of the model parameters. We observed a good mixing across most parameters. In particular, the average ESS was equal to 23,595. Moreover, more than 95% of the $\hat{R}$ values were below the standard threshold of 1.10. Posterior predictive checks based on sample mean and variance showed that observed values were consistent with the distribution of replicated data, supporting the adequacy of the model. Additionally, the 95% credible intervals included the observed value in all replications. In our analysis, we identified several significant non-linear effects. Figure 3 shows the MPPI for each main and interaction effect, both linear and non-linear, across 50 independent replicates. The significant main non-linear associations with BMI are VO2max, as shown in Figure 4, and C-reactive protein (crp). For linear associations, the significant relationships with BMI include insulin, displayed in Figure 4, VO2max, uric acid (sua), and gender.

As shown in Figure 3, we identified a linear interaction effect with MPPI of approximately 0.5. This effect involves the interaction between the variables VO2max and age, as illustrated in the left panel of Figure 5. Additionally, since we perform subject-specific inference, in Figure 5 we present the effect of age on the linear varying coefficient associated with VO2max. We observe an increasing trend of VO2max with age.

For the sake of comparison, we report the effects selected by the competing methods when possible. Specifically, the main effects selected by NLint are ins, cot, vomax, sua, pep, bap, mono, neno, albu, crp, ema, and emo. However, the model does not allow us to determine whether each effect was identified as linear or non-linear. The selected interaction effects are: ins × pep, vomax × pep, vomax × bap, pep × bap, ins × albu, pep × albu, mono × crp, vomax × emo, pep × ema, pep × emo, bap × ema, and ema × emo. As in the case of main effects, we do not know the nature of these interactions, whether they are linear or non-linear. In this case, discrete variables were included as control covariates; that is, the model does not perform variable selection on discrete covariates. The main effects selected by the BLASSO method are: ins, cot, vomax, sua, bap, crp, age, and male. The selected interaction effects include: ins × pep, vomax × pep, vomax × male, pep × age, pep × male, mono × crp, and age × male. In this case, since the BLASSO model selects only linear effects, all identified main and interaction effects are linear. The other methods, SVM, BART, and MLR, are not designed for variable selection and are therefore excluded from this part of the analysis.

6.2 Prediction of concrete compressive strength

The dataset used in this analysis was obtained from the UCI Machine Learning Repository (Yeh 2007) and focuses on Concrete Compressive Strength. It consists of 1,030 entries in the compressive strength experiment, each containing infor-

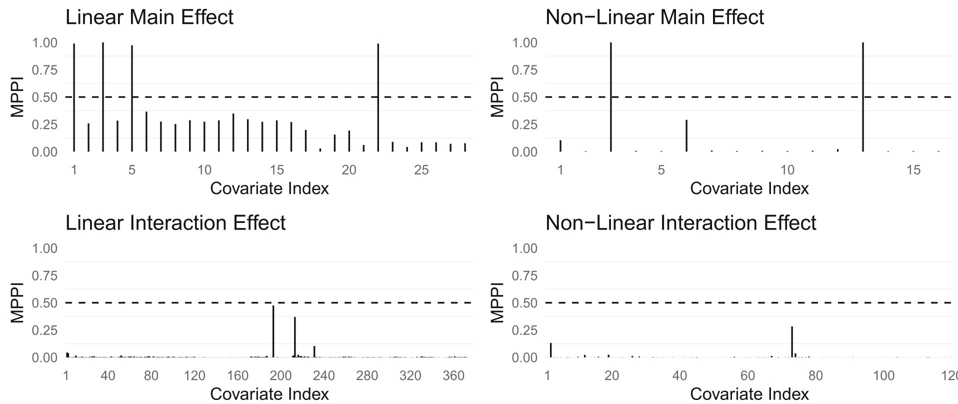

Fig. 3 Plot of MPPI for main and interaction effect both linear and non-linear, for the BMI dataset. The dashed line at 0.5 represents the median model threshold. The covariates are ordered according to the dataset, with all continuous variables first, followed by the categorical ones. Interactions are ordered in ascending column-wise order

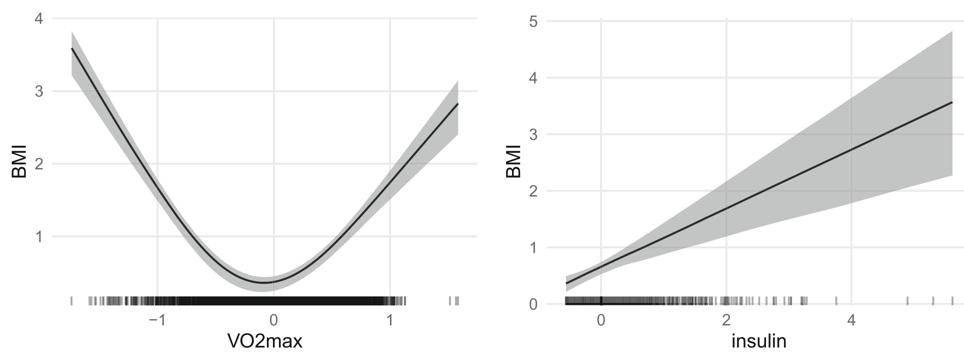

Fig. 4 Estimated response curves for the VO2max and insulin, when all the other quantities are equal to their median. The black line corresponds to the posterior median, the shaded bands indicate 95% posterior credible intervals and the marks on the x-axis indicate the observed data points

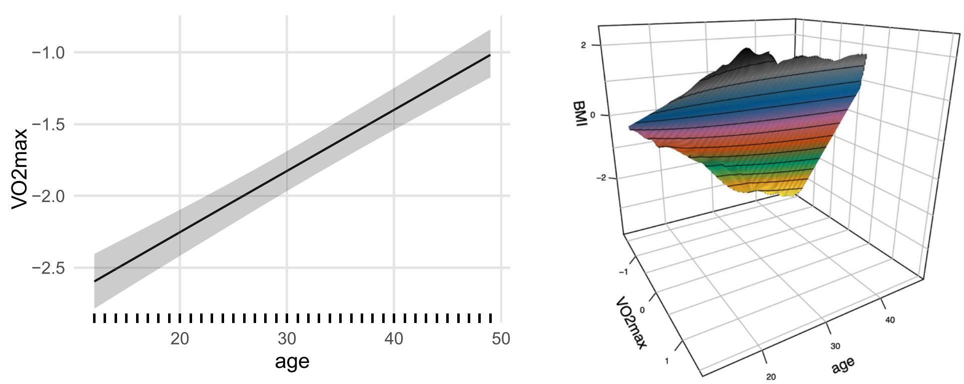

Fig. 5 *Left panel*: Estimated VO2max varying-coefficient curves associated with age, when all the other quantities are equal to their median. The black line corresponds to the posterior median, the shaded bands indicate 95% posterior credible intervals and the marks on the x-axis indicate the observed data points. *Right panel*: Estimated response surface for the interaction between VO2max and age, when all the other quantities are equal to their median

mation on eight different attributes. In particular, the data set contains no missing data or duplicate rows. The covariates include key ingredients such as cement and water, as well as the age of the concrete at the time of testing. The objective of this analysis is to predict the compressive strength based on these attributes.

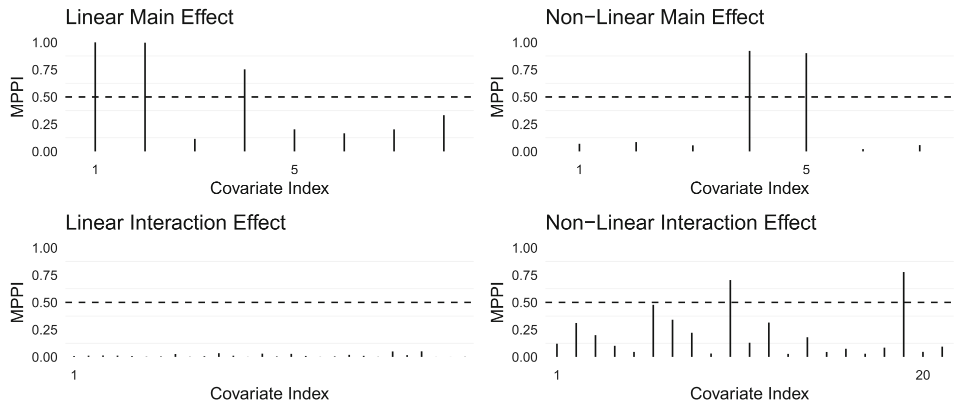

Fig. 6 Plot of MPPI for main and interaction effect both linear and non-linear, for the Concrete dataset. The dashed line at 0.5 represents the median probability model threshold. The covariates are ordered according to the dataset, with all continuous variables first, followed by the categorical ones. Interactions are ordered in ascending column-wise order

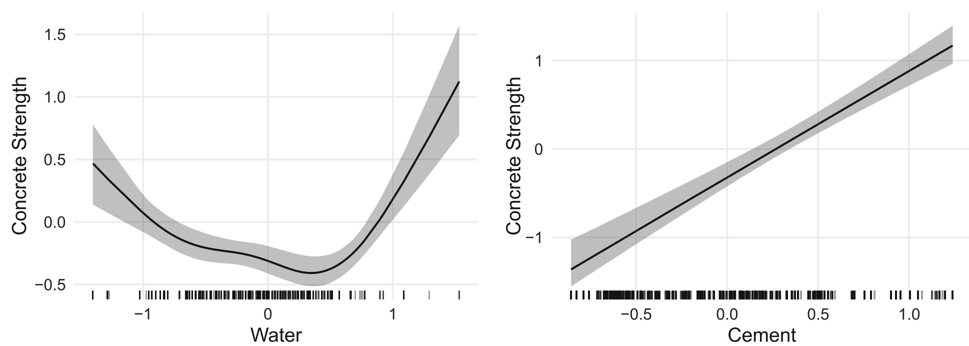

Fig. 7 Estimated response curves for Water and Cement, when all the other quantities are equal to their median. The black line corresponds to the posterior median, the shaded bands indicate 95% posterior credible intervals and the marks on the x-axis indicate the observed data points

Consistent with previous analysis, we executed 5 parallel MCMC chains, each with 500,000 iterations, a burn-in of 250,000, and thinning every 2 samples. Most of the parameters showed good mixing behavior, with the average ESS equal to 158,273. Furthermore, more than 95% of the $\hat{R}$ values were below the standard threshold of 1.1. A posterior predictive check was performed to assess the adequacy of the model. The results showed good agreement between the observed data and data simulated from the posterior predictive distribution. Specifically, a check was performed on the mean and variance of the sample, and in all replications, the observed values were within the corresponding 95% credible intervals.

The predictive results, shown in the second column of Table 2, reveal a strong non-linear structure in the data. Notably, methods like BART that are able to capture high-order interactions demonstrate better performance in modeling these complex relationships. Figure 6 illustrates the model's selection of three main linear effects, as well as two significant non-linear effects. It also highlights the selection of two non-linear interaction effects, while no linear interactions were included in the model. The significant main linear effects associated with Concrete Strength are Cement, Blast Furnace Slag, and Water. Additionally, significant non-linear associations were identified for Water and Superplasticizer. The significant main non-linear associations with Concrete Strength and Water, as shown in Figure 7. For linear associations, the significant relationships with Concrete Strength and Cement are displayed in Figure 7.

We identified a significant non-linear interaction effect between the variables Water × Fine Aggregate, and Water × Superplasticizer, as shown in Figure 9. In Figure 8 we present the effect of Superplasticizer on the non-linear varying coefficient associated with the CoarseAggregate and FineAggregate predictors.

The main effects selected by the NLint method are: Cement, BlastFurnaceSlag, FlyAsh, Water, and Superplasticizer. The interaction effects identified by NLint include: Cement × Water, Cement × Superplasticizer, FlyAsh × Superplasticizer, and Water × Superplasticizer. However, NLint does not allow us to determine whether the selected effects are linear or non-linear, limiting the interpretability of the results.

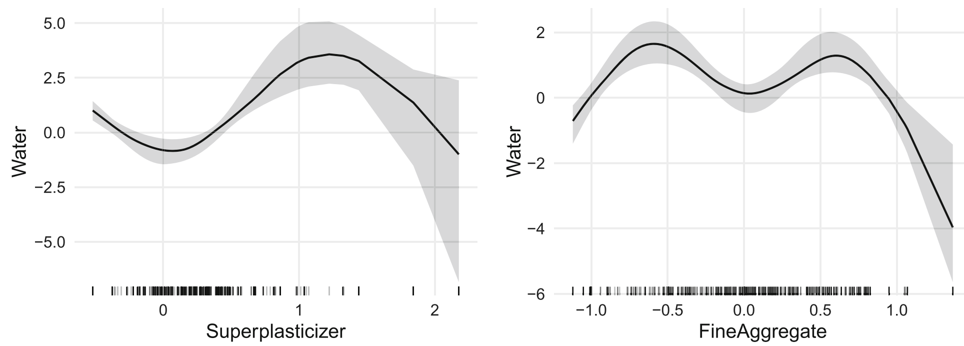

Fig. 8 Estimated Water varying-coefficient curves associated with Superplasticizer and Fine Aggregate, when all the other quantities are equal to their median. The black line corresponds to the posterior median, the shaded bands indicate 95% posterior credible intervals and the marks on the x-axis indicate the observed data points

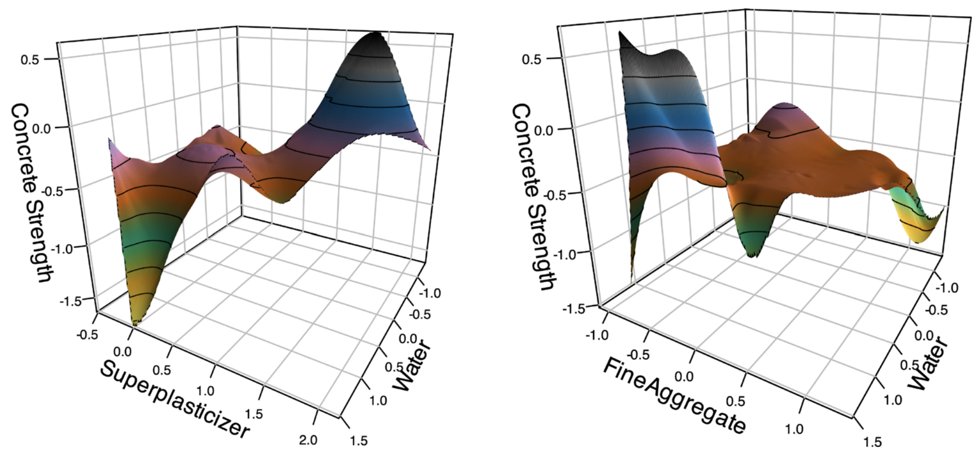

Fig. 9 Estimated response surface for the interactions between Fine Aggregate and Water, and Superplasticizer and Water, when all the other quantities are equal to their median

In contrast, the BLASSO method selects only Cement as a main effect and identifies no interaction effects. This result is consistent with the findings from our proposed model, which does not select any linear effects but identifies two non-linear effects that BLASSO is unable to capture due to its linear nature.

Table 2 Application prediction results: MSE mean across 50 different train-test splits (standard errors are in parentheses). The best predictive performance is in bold.

Method	BMI Dataset	Concrete Dataset
INVENT	**0.24** (0.03)	0.22 (0.04)
BART	0.27 (0.02)	**0.06** (0.01)
MLR	0.32 (0.01)	0.40 (0.03)
SVM	1.94 (2.46)	0.35 (0.05)
BLASSO	0.30 (0.03)	0.44 (0.04)
NLint	0.27 (0.02)	0.35 (0.04)
MixSelect	0.33 (0.04)	0.35 (0.04)
VCBART	0.25 (0.01)	0.21 (0.02)

7 Conclusions

We have proposed a flexible model for non-linear regression. Our model construction builds upon non-linear varying coefficients and ensures both flexibility and interpretability. In particular, covariates have a double role, acting both as predictors and effect modifiers; in practice, we need to specify an ordering of the covariates, but we demonstrated that the design matrix of the model is invariant to this ordering. We showed that varying coefficients lead to two-way non-linear interactions; the inclusion of all possible interactions results in a regression model with a large number of parameters, even with a limited number of covariates.

Given the large number of parameters, we introduced a spike-and-slab prior for both the main effects (linear and non-linear) and the interaction effects (linear and non-linear). This strategy allowed us to impose sparsity and select more interpretable models. Through simulation studies and two data analyses, we empirically demonstrated the ability of the model to detect non-linear dependencies and enable efficient variable selection in various scenarios, as well as its good prediction performance.

This emphasis on structural interpretability distinguishes our method from primarily predictive approaches such as BART or VCBART, the latter being discussed in Section 5 of the Supplementary Material. Although this paper focuses on a continuous outcome, the hierarchical structure defined to model complex associations can be adapted to response variables that belong to the exponential family. Furthermore, the model can be extended to include interactions of higher order, allowing interaction effects $\boldsymbol{\omega}$ to be covariate dependent.

The proposed approach is flexible, easily interpretable, but computationally demanding; nonetheless, our approach remains competitive in terms of execution time compared to competing methods. The choice of the heredity assumption affects the computational efficiency of the model; assuming strong heredity significantly reduces the computational cost compared to the other heredity assumptions.

Future research will consider strategies, such as adaptive MCMC approaches, that can accelerate the algorithm's convergence and result in reduced computational times; alternatively, variational methods may be a viable option for point estimation in high-dimensional settings. Additionally, it might be useful to consider shrinkage priors, such as the Horseshoe prior (Carvalho et al. 2010; Johndrow et al. 2020), as a less computationally demanding alternative to spike-and-slab priors.

Supplementary information

In this document, we provide the proofs for Theorem 1 and Corollary 1, demonstrating the invariance of the proposed method to the ordering of covariate indices, supported by empirical evidence (Section 1). We then detail the algorithm developed for posterior computation (Section 2). Furthermore, we report additional simulation studies (Section 3), analyze the sensitivity of the results to hyperparameter specifications (Section 4), and present additional comparisons with spikeSlabGAM, MixSelect, and VCBART methods (Section 5). (.pdf)

Code

An R package implementing the proposed method is available on GitHub at https://github.com/davidefabbrico/invent. The code consists of R and C++ functions integrated using the Rcpp and RcppArmadillo packages (Eddelbuettel and François 2011; Eddelbuettel and Sanderson 2014).

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s11222-026-10903-y.

Author Contributions D.F. and F.C.S. contributed to the development of the methodology. D.F., with the support of M.P., implemented the algorithm, conducted the simulation studies, and performed the real data analyses. D.F. also developed and proved the theoretical result included in the manuscript. M.P. and F.C.S. drafted the majority of the introduction, which was then revised and finalized by D.F. D.F. wrote the remaining sections of the manuscript. M.P. and F.C.S. provided critical revisions. F.C.S. supervised the project and contributed to the final version of the manuscript. All authors reviewed and approved the final manuscript.

Funding Open access funding provided by Università degli Studi di Firenze within the CRUI-CARE Agreement. The first and last authors were partially supported by the Italian Ministry of University and Research (MUR), Department of Excellence project 2023-2027 ReDS 'Rethinking Data Science' - Department of Statistics, Computer Science, Applications - University of Florence.
The last author was partially supported by the European Union - NextGenerationEU - National Recovery and Resilience Plan, Mission 4 Component 2 - Investment 1.5 - THE - Tuscany Health Ecosystem - ECS00000017 - CUP B83C22003920001 and the MUR-PRIN grant 2022 SMNNKY, CUP B53D23009470006, funded by the European Union Next Generation EU, Mission 4, Component 2.

Data Availability The data used for the analyses are publicly available online. Specifically, the dataset for the BMI analysis can be accessed from the NHANES (National Health and Nutrition Examination Survey) website at: https://wwwn.cdc.gov/nchs/nhanes/Default.aspx. The data related to concrete compressive strength are available from the UCI Machine Learning Repository at the following link:https://archive.ics.uci.edu/dataset/165/concrete+compressive+strength.

Declarations

Competing interests The authors declare no competing interests.

References

Antonelli, J., Mazumdar, M., Bellinger, D., Christiani, D., Wright, R., Coull, B.: Estimating the health effects of environmental mixtures

using bayesian semiparametric regression and sparsity inducing priors, (2020)
Barbieri, M.M., Berger, J.O.: Optimal predictive model selection, (2004)
Bien, J., Taylor, J., Tibshirani, R.: A lasso for hierarchical interactions. Ann. Stat. **41**(3), 1111 (2013)
Chipman, H.A., George, E.I., McCulloch, R.E.: Bart: Bayesian additive regression trees (2010)
Chipman, H., McCulloch, R., Chipman, G.: Package ‘BayesTree’ (2016)
Cordell, H.J.: Detecting gene-gene interactions that underlie human diseases. Nat. Rev. Genet. **10**(6), 392–404 (2009)
Carvalho, C.M., Polson, N.G., Scott, J.G.: The horseshoe estimator for sparse signals. Biometrika **97**, 465–480 (2010)
Deshpande, S.K., Bai, R., Balocchi, C., Starling, J.E., Weiss, J.: VCBART: Bayesian Trees for Varying Coefficients. Bayesian Anal. pp. 1–28 (2024). https://doi.org/10.1214/24-BA1470
Eddelbuettel, D., François, R.: Rcpp: Seamless r and c++ integration. J. Stat. Softw. **40**, 1–18 (2011)
Eddelbuettel, D., Sanderson, C.: Rcpparmadillo: Accelerating r with high-performance c++ linear algebra. Comput. Stat. Data Anal. **71**, 1054–1063 (2014)
Ferrari, F., Dunson, D.B.: Identifying main effects and interactions among exposures using gaussian processes. Ann. Appl. Stat. **14**(4), 1743 (2020)
Gustafson, P.: Bayesian regression modeling with interactions and smooth effects. J. Am. Stat. Assoc. **95**(451), 795–806 (2000)
Gelman, A., Van Dyk, D.A., Huang, Z., Boscardin, J.W.: Using redundant parameterizations to fit hierarchical models. J. Comput. Graph. Stat. **17**(1), 95–122 (2008)
Hearst, M.A., Dumais, S.T., Osuna, E., Platt, J., Scholkopf, B.: Support vector machines. IEEE Intell. Syst. Appl. **13**(4), 18–28 (1998)
Hastie, T., Tibshirani, R.: Varying-coefficient models. J. Roy. Stat. Soc.: Ser. B (Methodol.) **55**(4), 757–779 (1993)
Im, Y., Huang, Y., Tan, A., Ma, S.: Bayesian finite mixture of regression analysis for cancer based on histopathological imaging-environment interactions. Biostatistics **24**(2), 425–442 (2023)
Johndrow, J.E., Orenstein, P., Bhattacharya, A.: Scalable approximate MCMC algorithms for the Horseshoe prior. J. Mach. Learn. Res. **21**, 73–17361 (2020)
Liu, S.H., Bobb, J.F., Lee, K.H., Gennings, C., Claus Henn, B., Bellinger, D., Austin, C., Schnaas, L., Tellez-Rojo, M.M., Hu, H., et al.: Lagged kernel machine regression for identifying time windows of susceptibility to exposures of complex mixtures. Biostatistics **19**(3), 325–341 (2018)
Lim, M., Hastie, T.: Learning interactions via hierarchical group-lasso regularization. J. Comput. Graph. Stat. **24**(3), 627–654 (2015)
Lampa, E., Lind, L., Lind, P.M., Bornefalk-Hermansson, A.: The identification of complex interactions in epidemiology and toxicology: a simulation study of boosted regression trees. Environ. Health **13**(1), 1–17 (2014)
Marchini, J., Donnelly, P., Cardon, L.R.: Genome-wide strategies for detecting multiple loci that influence complex diseases. Nat. Genet. **37**(4), 413–417 (2005)
Ni, Y., Stingo, F.C., Baladandayuthapani, V.: Bayesian graphical regression. J. Am. Stat. Assoc. **114**(525), 184–197 (2019)
Pedone, M., Amedei, A., Stingo, F.C.: Subject-specific dirichlet-multinomial regression for multi-district microbiota data analysis. Ann. Appl. Stat. **17**(1), 539–559 (2023)
Papadogeorgou, G., Bello, C., Ovaskainen, O., Dunson, D.B.: Covariate-informed latent interaction models: Addressing geographic & taxonomic bias in predicting bird-plant interactions. J. Am. Stat. Assoc. pp. 1–12 (2023)
Radchenko, P., James, G.M.: Variable selection using adaptive nonlinear interaction structures in high dimensions. J. Am. Stat. Assoc. **105**(492), 1541–1553 (2010)
Ruppert, D., Wand, M.P., Carroll, R.J.: Semiparametric Regression, vol. 12. Cambridge University Press, Cambridge (2003)
Scott, J.G., Berger, J.O.: Bayes and empirical-bayes multiplicity adjustment in the variable-selection problem. Ann. Stat. pp. 2587–2619 (2010)
Scheipl, F.: spikeslabgam: Bayesian variable selection, model choice and regularization for generalized additive mixed models in r. J. Stat. Softw. **43**, 1–24 (2011)
Stingo, F.C., Chen, Y.A., Tadesse, M.G., Vannucci, M.: Incorporating biological information into linear models: A bayesian approach to the selection of pathways and genes. Ann. Appl. Stat. **5**(3), 10–1214 (2011)
Scheipl, F., Fahrmeir, L., Kneib, T.: Spike-and-slab priors for function selection in structured additive regression models. J. Am. Stat. Assoc. **107**(500), 1518–1532 (2012)
Savitsky, T., Vannucci, M., Sha, N.: Variable selection for nonparametric gaussian process priors: Models and computational strategies. Stat. Sci: Rev J. Inst. Math. Stat. **26**(1), 130 (2011)
Xu, X., Ghosh, M.: Bayesian variable selection and estimation for group lasso, (2015)
Yeh, I.-C.: Concrete Compressive Strength. UCI Machine Learning Repository. https://doi.org/10.24432/C5PK67 (2007)
Zhao, P., Rocha, G., Yu, B.: The composite absolute penalties family for grouped and hierarchical variable selection. Ann. Stat. **37**(6A), 3468–3497 (2009). https://doi.org/10.1214/07-AOS584